AD618112

BOUNDS FOR LATTICE DISTRIBUTIONS
HAVING MONOTONE HAZARD RATE
WITH APPLICATIONS

by

Gamanlal P. Shah

OPERATIONS RESEARCH CENTER

COLLEGE OF ENGINEERING
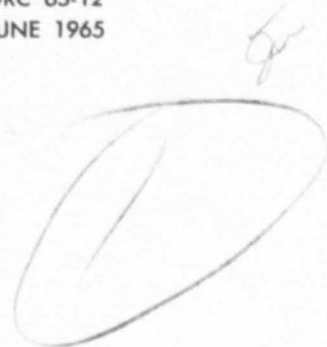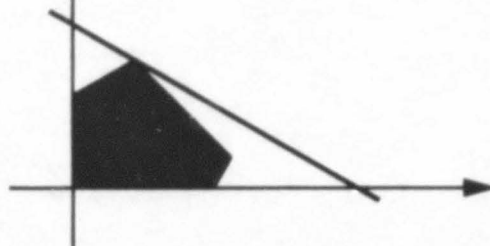
UNIVERSITY OF CALIFORNIA - BERKELEY

Bounds for Lattice Distributions Having Monotone

Hazard Rate With Applications

by

Gamanlal P. Shah
Operations Research Center
University of California, Berkeley

June 1965

ORC 65-12

Summary

Using methods which have been employed for obtaining bounds on continuous failure distributions, bounds on lattice distributions are derived under the increasing (decreasing) failure rate assumptions. The discrete bounds are convenient in a number of applications such as life table analysis and reliability theory.

A

1. Introduction

A discrete distribution of a random variable X is called a lattice distribution if there exist real numbers a and $h > 0$ such that every possible value of X can be represented in the form $a + kh$, where k runs through integer values. We call h the span of the distribution. The geometric, binomial and Poisson distributions are important examples of lattice distributions.

It seems that there is essentially no literature on Chebyshev type bounds for lattice distributions. Of course, the classical Chebyshev bounds are attained by discrete distributions. However, these need not be lattice. The general method used by Karlin and Studden and others [12, 11] for obtaining Chebyshev type bounds can, of course, in theory be applied to lattice distributions. However, we will restrict ourselves to a special class of lattice distributions where these methods do not provide sharp bounds.

For convenience assume that $a = 0$, $h = 1$ and let $p_k = P[X = k]$; $(k = 0,1,2,\ldots)$. Define the failure rate $r(k)$ for lattice distributions as

$$r(k) = \frac{p_k}{\displaystyle\sum_{j=k}^{\infty} p_j} \quad \text{if} \quad \sum_{j=k}^{\infty} p_j > 0$$

and note that $0 \leq r(k) \leq 1$. Let $F(k) = \displaystyle\sum_{j=0}^{k-1} p_j$ and $\bar{F}(k) = 1 - F(k) = \displaystyle\sum_{j=k}^{\infty} p_j$. We say that F is lattice IFR (lattice DFR)

if and only if $r(k)$ is non-decreasing (non-increasing) in k .

Note that it is shown in [3] that F is lattice IFR (lattice DFR)

iff $\log \overline{F}(k)$ is concave (convex). Further equivalent definitions

of lattice IFR (lattice DFR) distributions in terms of Pólya

frequency function of order two and total positivity of order two have

been discussed in [1,2,3,10]. Note that the negative binomial

distribution

$$P_k = \binom{k-1}{\alpha-1} p^\alpha (1-p)^{k-\alpha} \quad ; \quad \begin{array}{c} \alpha > 0 \\ k = 0,1,2,\ldots. \end{array}$$

is lattice IFR (increasing failure rate) for $\alpha > 1$ and lattice DFR

(decreasing failure rate) for $\alpha < 1$. For $\alpha = 1$, it is the

geometric distribution with a constant failure rate and hence it

belongs to both IFR and DFR classes. The geometric distribution

is the only distribution with this property in the discrete case.

The binomial and Poisson distributions are lattice IFR.

Lattice distributions occur naturally in many applications.

Fatigue failure data obtained under dynamic loading is commonly

recorded in terms of the number of cycles to failure. For many

structures subject to fatigue failure, a lattice IFR distribution

would seem appropriate. A mathematical model describing such types

of failure has been discussed in [5]. As another example we mention

human mortality data where it is common to record time of death

to the nearest year or the nearest month. For certain types of

cancer mortality data a decreasing failure rate seems appropriate, [4].

In any situation where grouped data is derived from a continuous

IFR(DFR) distribution, a lattice IFR (lattice DFR) distribution will be

appropriate.

2. Bounds on Lattice Distributions

2.1 IFR Bound:

Let F denote a left continuous distribution function such that $F(0^-) = 0$. Bounds on $1-F$ under the IFR (DFR) assumption in the continuous case when the first moment μ_1 is given can be found in [2]. In the IFR case, the lower bound is given by Theorem 1.

Theorem 1: Let F be a left continuous distribution. Let F be IFR with mean μ_1, then

$$\bar{F}(t) \geq \begin{cases} e^{-\frac{t}{\mu_1}} & ; \quad t < \mu_1 \\ 0 & ; \quad t \geq \mu_1 \end{cases}.$$

The inequality is sharp.

A similar bound exists for lattice distributions and we shall discuss this after proving the following lemma.

Lemma: Let F be a lattice distribution with mean μ_1. Assume that F is IFR and let $s_k = \sum_{i=1}^{k} \bar{F}(i)$ then for $t \leq \mu_1$

$$(1) \qquad \bar{F}(t) \geq \bar{F}(t-k) \left(1 - \frac{\bar{F}(t-k)}{\mu_1 - s_{t-k-1}} \right)^k ; \quad k = 1, 2, \ldots, t-1.$$

Proof: We shall prove (1) by induction. By definition, $\bar{F}(k) = \sum_{i=k}^{\infty} p_i$, so that $\log \bar{F}(k) = \log \left(\sum_{i=k}^{\infty} p_i \right)$. Note that the function $\log \bar{F}(k)$ is defined only for integer values. By linear interpolation we can define $\log \bar{F}(k)$ for any real value of k as shown in Figure 2.1.

Since we assume that F is IFR, it follows that $\log \overline{\overline{F}}(k)$ is concave. Hence there exists a supporting line to $\log \overline{F}(k)$ as shown in Figure 2.1.
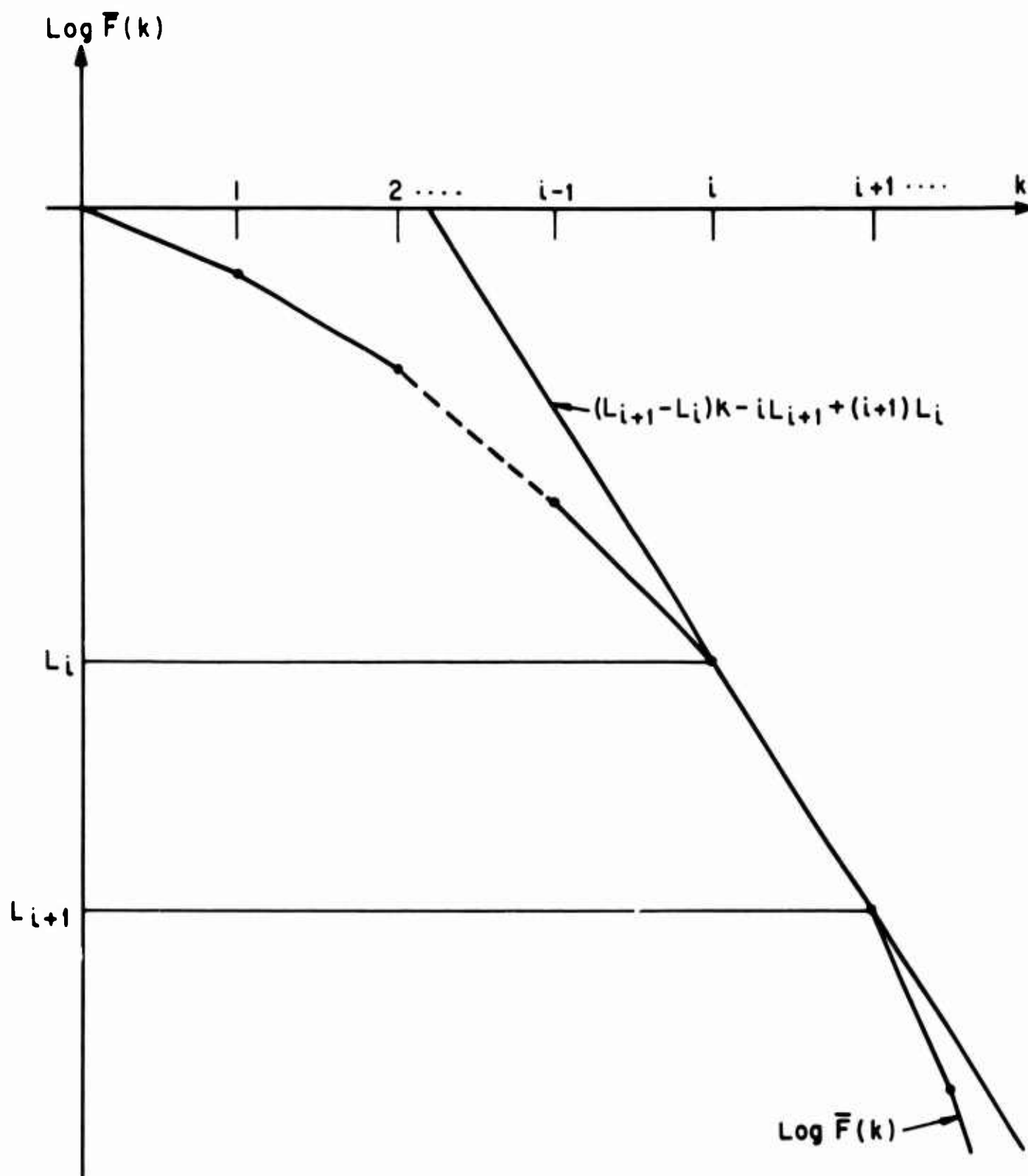
FIGURE 2.1 SHOWING SUPPORTING LINE FOR LOG $\overline{F}(k)$

Thus by log concavity, we have

$$\log \overline{F}(k) \leq (L_{i+1} - L_i)k - iL_{i+1} + (i+1)L_i \; ; \; k = i, i+1, \ldots \;.$$

But $L_i = \log \overline{F}(i)$ and $L_{i+1} = \log \overline{F}(i+1)$, therefore

$$\overline{F}(k) \leq \left[\frac{\overline{F}(i+1)}{\overline{F}(i)} \right]^k \frac{[\overline{F}(i)]^{i+1}}{[\overline{F}(i+1)]^i} \quad \text{for} \quad k = i, i+1, \ldots$$

Let $s_k = \overline{F}(1) + \overline{F}(2) + \ldots + \overline{F}(k)$ where $s_0 \equiv 0$. Then we have

$$\mu_1 = \sum_{k=1}^{\infty} \overline{F}(k) = \sum_{k=1}^{i-1} \overline{F}(k) + \sum_{k=i}^{\infty} \overline{F}(k)$$

$$\leq s_{i-1} + \sum_{k=i}^{\infty} \left[\frac{\overline{F}(i+1)}{\overline{F}(i)} \right]^k \frac{[\overline{F}(i)]^{i+1}}{[\overline{F}(i+1)]^i}$$

$$= s_{i-1} + \overline{F}(i) \sum_{k=i}^{\infty} \left[\frac{\overline{F}(i+1)}{\overline{F}(i)} \right]^{k-i} \;.$$

Thus $\mu_1 \leq s_{i-1} + \dfrac{\overline{F}(i)^2}{\overline{F}(i) - \overline{F}(i+1)}$. Solving for $\overline{F}(i+1)$, we obtain

$$(2) \qquad \overline{F}(i+1) \geq \overline{F}(i) - \frac{\overline{F}(i)^2}{\mu_1 - s_{i-1}}$$

From (2) with $i = t-1$, it follows that (1) holds for $k = 1$. Now assume that (1) is true for $k = n-1$; $(n \leq t)$ i.e.;

$$(3) \qquad \overline{F}(t) \geq \overline{F}(t-n+1) \left(1 - \frac{\overline{F}(t-n+1)}{\mu_1 - s_{t-n}} \right)^{n-1} = \emptyset(\overline{F}(t-n+1)$$

where $\emptyset(x) = x \left(1 - \dfrac{x}{\mu_1 - s_{t-n}} \right)^{n-1}$.

Note that $\phi'(x) = \left(1 - \dfrac{x}{\mu_1 - s_{t-n}}\right)^{n-2} \left(1 - \dfrac{nx}{\mu_1 - s_{t-n}}\right)$

$$\geq 0 \quad \text{if} \quad x \leq \frac{\mu_1 - s_{t-n}}{n} \quad .$$

<u>To Show</u>: $x \leq \dfrac{\mu_1 - s_{t-n}}{n}$

By definition $x = \overline{F}(t-n+1) \leq 1$, hence we need only show $\dfrac{\mu_1 - s_{t-n}}{n} \geq 1$.
But $s_{t-n} \leq t-n$ implies $\dfrac{\mu_1 - s_{t-n}}{n} \geq \dfrac{\mu_1 - t + n}{n} \geq 1$ since $t \leq \mu_1$
by assumption.

Thus we can substitute for $\overline{F}(t-n+1)$ in (3) the bound given
by (2) with $i = t-n$ and obtain

$$\overline{F}(t) \geq \overline{F}(t-n) \left(1 - \frac{\overline{F}(t-n)}{\mu_1 - s_{t-n-1}}\right) \left[1 - \frac{\overline{F}(t-n)\left(1 - \frac{\overline{F}(t-n)}{\mu_1 - s_{t-n-1}}\right)}{\mu_1 - s_{t-n}}\right]^{n-1}$$

Noting $s_{t-n} = s_{t-n-1} + \overline{F}(t-n)$ and letting $c = 1 - \dfrac{\overline{F}(t-n)}{\mu_1 - s_{t-n-1}}$,
the above inequality becomes

$$\overline{F}(t) \geq \overline{F}(t-n) \, c \left[\frac{\left(1 - \frac{\overline{F}(t-n)}{\mu_1 - s_{t-n-1}}\right)^2}{\frac{\mu_1 - s_{t-n}}{\mu_1 - s_{t-n-1}}}\right]^{n-1}$$

$$= \overline{F}(t-n) \, c \left[\frac{c^2}{c}\right]^{n-1} = c^n \overline{F}(t-n) \quad .$$

This is (1) for $k = n$, and the induction is complete.

<u>Theorem 2</u>: If $F(t)$ is a lattice IFR distribution with mean
$\mu_1 \geq 1$ placing mass p_k at $k = 0,1,2,\ldots$ and $\overline{F}(t) = \displaystyle\sum_{k=t}^{\infty} p_k$; then

-7-

$$\bar{F}(t) \geq \left[\begin{array}{ll} \left(\dfrac{\mu_1}{1+\mu_1}\right)^t & \text{for } t \leq \mu_1 \\[3mm] \dfrac{\mu_1 - [\mu_1]}{1+\mu_1 - [\mu_1]} & \text{for } t = [\mu_1] + 1 \\[3mm] 0 & \text{for } t > [\mu_1] + 1 \end{array} \right.$$

Where $[\mu_1]$ denotes the greatest integer contained in μ_1. The inequality is sharp.

Proof: Since $\log \bar{F}(k)$ is concave in k and noting figure 2.1 for $i = 0$, we see that

$$\log \bar{F}(k) \leq L_1 k$$
or $$\bar{F}(k) \leq [\bar{F}(1)]^k$$

Thus $\mu_1 = \displaystyle\sum_{k=1}^{\infty} \bar{F}(k) \leq \sum_{k=1}^{\infty} [\bar{F}(1)]^k = \dfrac{\bar{F}(1)}{1-\bar{F}(1)}$

(4) or $\bar{F}(1) \geq \dfrac{\mu_1}{1+\mu_1}$

Now consider (1) with $k = t-1$; this gives

$$\bar{F}(t) \geq \bar{F}(1) \left[1 - \frac{\bar{F}(1)}{\mu_1} \right]^{t-1} = \psi(\bar{F}(1)) \quad .$$

Note that $\psi'(x) = (1 - \frac{x}{\mu_1})^{t-2}(1 - \frac{tx}{\mu_1}) \geq 0$ if $t \leq \mu_1$ and $\mu_1 \geq 1$ (since $x \leq 1$), and thus from (4), we have

$$\bar{F}(t) \geq \psi(\frac{\mu_1}{1+\mu_1}) = \left(\frac{\mu_1}{1+\mu_1} \right)^t \qquad \text{for} \quad t \leq \mu_1 \quad .$$

This bound for $t \leq \mu_1$ is attained by the geometric distribution; i.e.,

$$p_k = \frac{1}{1+\mu_1} \left(\frac{\mu_1}{1+\mu_1} \right)^k \qquad ; \quad k = 0,1,2,\ldots \ .$$

We conjecture that

$$\bar{F}(t) \geq \begin{cases} \dfrac{\mu_1 - [\mu_1]}{1+\mu_1 - [\mu_1]} & \text{for} \quad t = [\mu_1] + 1 \\[2em] 0 & \text{for} \quad t > [\mu_1] + 1 \end{cases}$$

where $[\mu_1]$ denotes the greatest integer contained in μ_1.

For $t > [\mu_1] + 1$, it is easily seen that the bound is attained by

$$\bar{G}(t) = \begin{cases} 1 & ; \quad t \leq [\mu_1] \\[1em] \dfrac{\mu_1 - [\mu_1]}{1 + [\mu_1]} & ; \quad t = [\mu_1] + 1 \\[1em] 0 & ; \quad t > [\mu_1] + 1 \end{cases}$$

The upper bound on $1-F$ for the continuous distribution when F is IFR is given by the following theorem.

Theorem 3:   Let F denote a left continuous distribution. Assume that F is IFR with mean μ_1, then

$$\bar{F}(t) \leq \begin{cases} 1 & ; \quad t \leq \mu_1 \\ e^{-wt} & ; \quad t > \mu_1 \end{cases}$$

where w depends on t and satisfies $1 - w\mu_1 = e^{-wt}$. The inequality is sharp [2].

An analogous bound exists for lattice distributions and is given by

<u>Theorem 4</u>:   Let F denote a lattice distribution. Assume that

F is IFR with mean μ_1 , then

$$\bar{F}(t) \leq \begin{cases} 1 & ; \quad t \leq \mu_1 \\ q^t & ; \quad t > \mu_1 \end{cases}$$

where q depends on t and satisfies $q\left(\dfrac{1-q^t}{1-q}\right) = \mu_1$. The inequality

is sharp.

<u>Proof</u>:   Let us define,

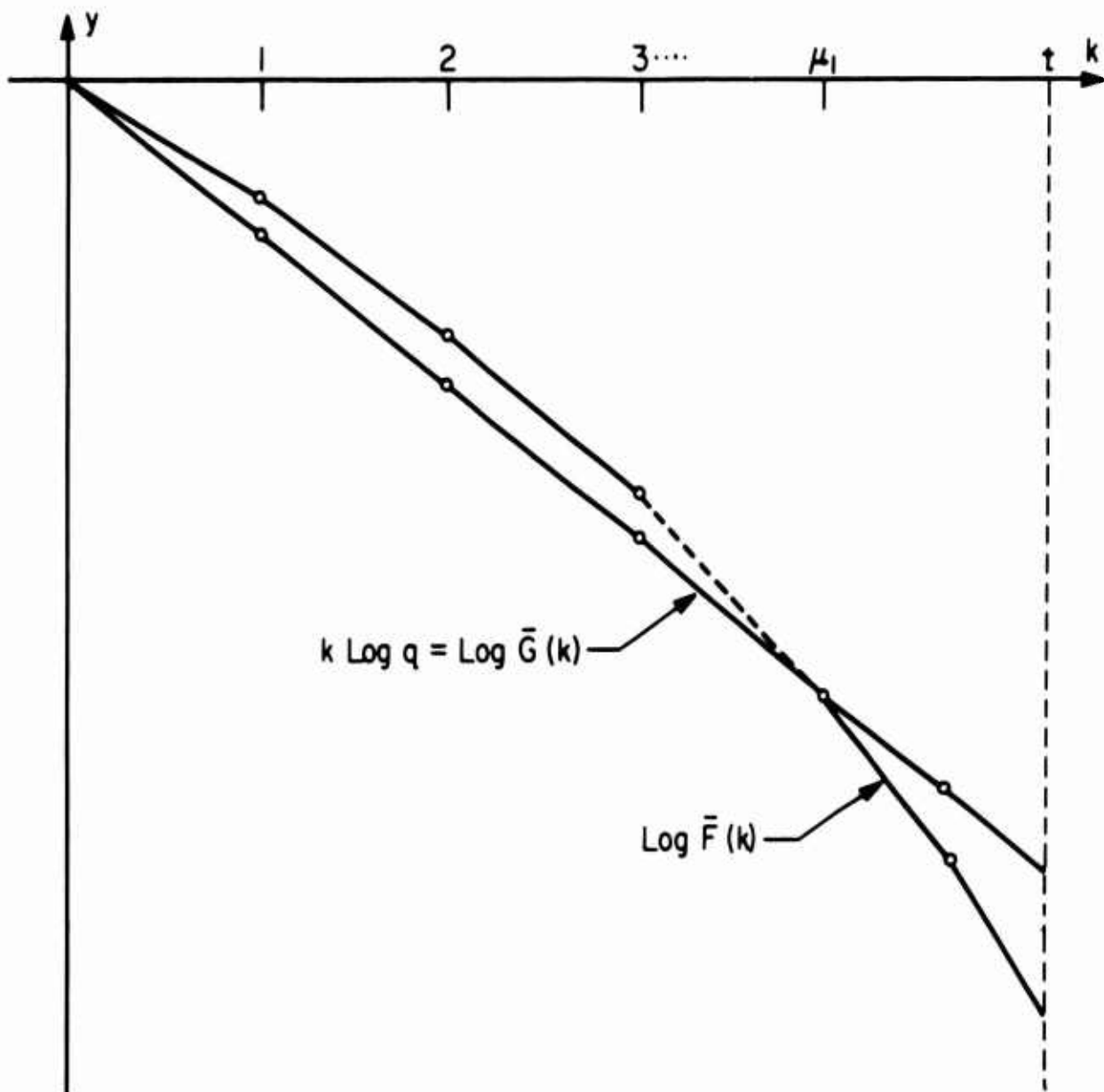$$\bar{G}(k) = \begin{cases} q^k & ; \quad k = 0,1,2,\dots,t \\ 0 & ; \quad k > t \ . \end{cases}$$

FIGURE 2.2 LOGARITHMIC CURVE OF THE FUNCTIONS $\bar{F}(k)$ AND $\bar{G}(k)$

Note that $\log \overline{G}(k)$ is linear and $\log \overline{F}(k)$ is concave, hence $\overline{F}(k)$ can cross $\overline{G}(k)$ at most once for $k < t$.

If $t > \mu_1$, we can always determine q so that

$$\sum_{k=1}^{\infty} \overline{G}(k) = \mu_1$$

or

$$\sum_{k=1}^{t} q^k = \mu_1$$

or

$$q(\frac{1-q^t}{1-q}) = \mu_1 \quad \text{which is solvable for } q.$$

Since, $\overline{F}(k)$ and $\overline{G}(k)$ are distribution functions with the same mean μ_1, they must cross at least once for $0 < k \leq t$.

Thus, for this choice of q, $\overline{F}(k)$ necessarily crosses q^k exactly once from above for $0 < k < t$. Therefore, $\overline{F}(t) < q^t$ unless F coincides identically with G. The degenerate distribution concentrating at μ_1 provides the upper bound for $t \leq \mu_1$ when μ_1 is an integer.

Since G is also IFR, the bound is sharp.

2.2 DFR Bound:

It is shown in [2] that the sharp lower bound on $1-F$ where F is a left continuous DFR distribution with known mean, say μ_1, is zero and the sharp upper bound is given by the following theorem.

Theorem 5: Let F be left continuous. Assume that F is DFR with mean μ_1, then

$$
\overline{F}(t) \leq
\begin{cases}
e^{-\frac{t}{\mu_1}} & ; \quad t < \mu_1 \\[2mm]
\dfrac{\mu_1 e^{-1}}{t} & ; \quad t \geq \mu_1
\end{cases}
$$

The inequality is sharp.

Similarly, for lattice distributions the sharp lower bound on 1-F where F is a lattice DFR distribution with known mean μ_1 is zero. For example, consider

$$
\overline{G}(k) =
\begin{cases}
0 & , \quad k < 0 \\[2mm]
a q^k & , \quad k = 0, 1, 2, \ldots
\end{cases}
$$

where a is arbitrarily small and $0 \leq q \leq 1$. Choose q such that G has mean μ_1. Then G is DFR with a jump at the origin and $\overline{G}(k) \leq a$ for all $k = 0, 1, 2, \ldots$. Letting $a \to 0$, we see that the sharp lower bound is zero.

Thus, for lattice DFR distributions only the upper bound on $\overline{F}(t)$ is given.

Theorem 6: Let F be a lattice distribution. Assume that F is DFR with mean μ_1, then

$$
\overline{F}(t) \leq
\begin{cases}
\left(\dfrac{\mu_1}{\mu_1 + 1}\right)^t & ; \quad t < \mu_1 \\[3mm]
\left(\dfrac{t-1}{t}\right)^t \left(\dfrac{\mu_1}{t-1}\right); & t \geq \mu_1
\end{cases}
$$

The inequality is sharp.

Proof: By definition, $\overline{F}(k) = \sum_{j=k}^{\infty} p_j$.

We remark that the function $\log \overline{F}(k)$ is defined only for integer values. By linear interpolation, we can define $\log \overline{F}(k)$ for any real value of k as shown in figure 2.3.

Since we assume that F is DFR, it follows that $\log \overline{F}(k)$ is convex, and therefore, there exists a supporting line to $\log \overline{F}(k)$, say, at k = t .
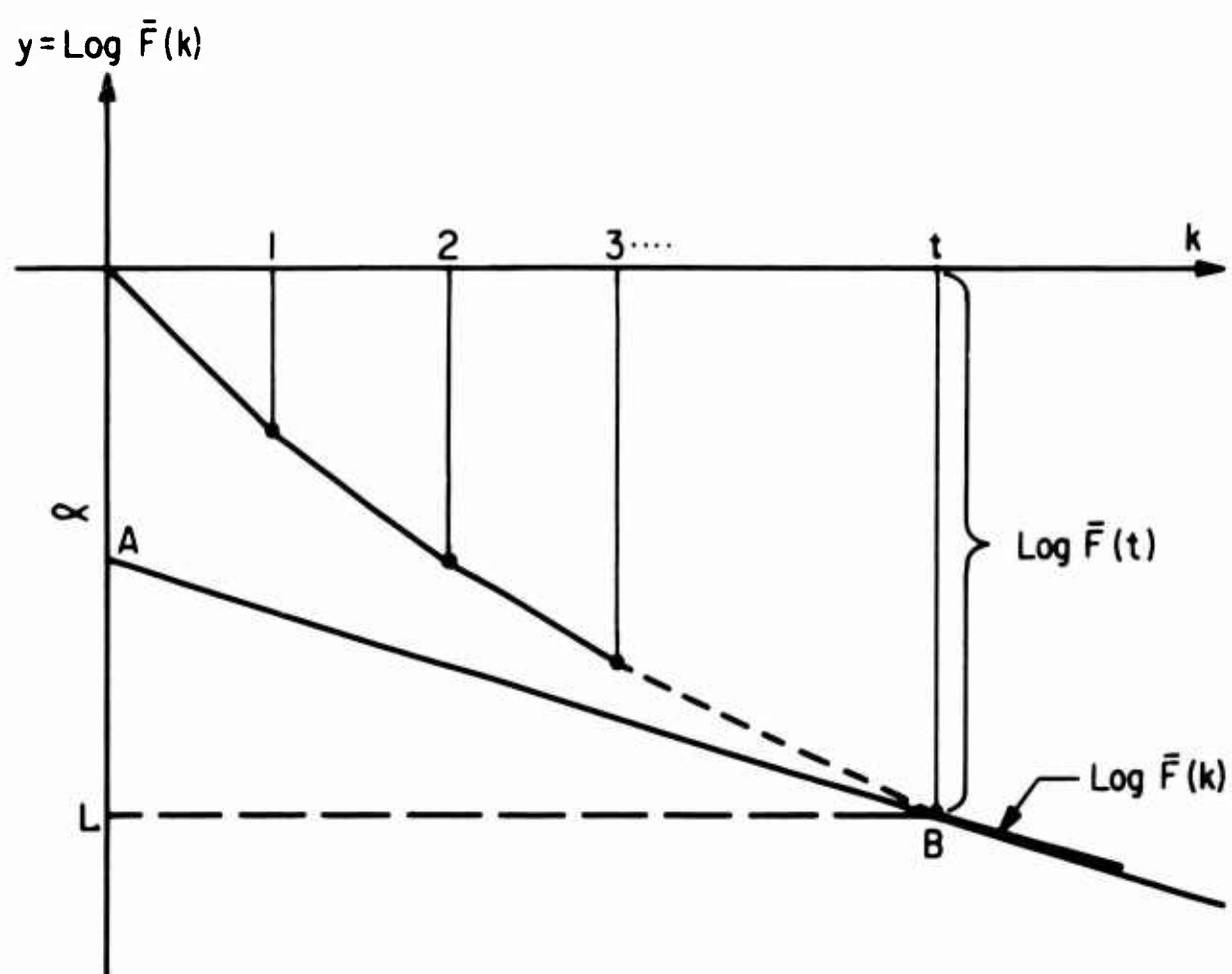
FIGURE 2.3 SHOWING SUPPORTING LINE FOR DFR DISTRIBUTION

Let α denote the intersection of the supporting line with y-axis and let $L = \log \bar{F}(t)$. Obviously, $L < \alpha \leq 0$ and the equation of the supporting line is $y = (\frac{L-\alpha}{t})k + \alpha$. Thus we have,

(5) $\qquad \log \bar{F}(k) \geq (\frac{L-\alpha}{t})k + \alpha \quad ; \quad k \geq 0$

or $\qquad \bar{F}(k) \geq e^{(\frac{L-\alpha}{t})k+\alpha}$.

Hence, $\qquad \mu_1 \geq \sum_{k=1}^{\infty} e^{ak+\alpha} \qquad$ where $\qquad a = \frac{L-\alpha}{t}$

$\qquad = e^{\alpha+a} \sum_{k=1}^{\infty} e^{a(k-1)}$

$\qquad = \dfrac{e^{\alpha+a}}{1-e^a} \qquad\qquad$ since $\qquad |e^a| < 1$.

That is

$$\mu_1 \geq \frac{e^{\alpha}}{e^{-a}-1} \qquad .$$

Solving for L, we obtain

$$L < \alpha - t \log(1 + \frac{e^{\alpha}}{\mu_1}) .$$

Let $\phi(\alpha) = \alpha - t \log(1 + \frac{e^{\alpha}}{\mu_1})$ so that

(6) $\qquad L \leq \phi(\alpha) \qquad\qquad$ for some α, $L < \alpha \leq 0$,

or $\qquad L \leq \max_{L < \alpha \leq 0} \phi(\alpha)$.

-16-

Note that

$$\phi'(\alpha) = 1 - \frac{t}{e^{-\alpha}\mu_1 + 1}$$

and

$$\phi''(\alpha) = \frac{-t\mu_1 e^{-\alpha}}{[e^{-\alpha}\mu_1 + 1]^2} < 0 \ .$$

Since $\phi''(\alpha) < 0$, we know that $\phi(\alpha)$ is strictly concave and hence it has a maxima, say, at $\alpha = \alpha*$.

Thus, from (6) we have,

$$(7) \qquad L \leq \phi(\alpha*)$$

so it remains to obtain $\alpha*$. $\alpha*$ can be found by equating $\phi'(\alpha)$ to zero; that is

$$1 - \frac{t}{e^{-\alpha}\mu_1 + 1} = 0 \ .$$

This gives us,

$$\alpha* = \log\left(\frac{\mu_1}{t-1}\right) \qquad \text{for} \ \ t \geq \mu_1 \ .$$

Note that for $t < \mu_1$,

$$\phi'(0) = 1 - \frac{t}{\mu_1 + 1} > 0 \ .$$

Also we have

$$\phi''(\alpha) < 0 \qquad \text{for any} \ \alpha \ \text{where} \ L < \alpha \leq 0 \ .$$

-17-

Thus, for $t < \mu_1$, we have $\phi'(\alpha) > 0$ and $\phi''(\alpha) < 0$ which shows that ρ is strictly increasing concave function and hence the maximum of $\phi(\alpha)$ is achieved at the origin.

That is

$$(8) \qquad \alpha^* = \begin{cases} 0 & \text{if } t < \mu_1 \\[2ex] \log\left(\dfrac{\mu_1}{t-1}\right) & \text{if } t \geq \mu_1 \ . \end{cases}$$

Using (8), the expression in (7) becomes,

$$L \leq \begin{cases} \phi(0) & \text{if } t < \mu_1 \\[2ex] \phi[\log(\dfrac{\mu_1}{t-1})] & \text{if } t \geq \mu_1 \end{cases}$$

or

$$L \leq \begin{cases} \log\left(\dfrac{\mu_1}{\mu_1+1}\right)^t & \text{if } t < \mu_1 \\[2ex] \log[(\dfrac{\mu_1}{t-1})(\dfrac{t-1}{t})^t] & \text{if } t \geq \mu_1 \ . \end{cases}$$

But $L = \log \bar{F}(t)$, so finally we have

$$(9) \qquad \bar{F}(t) \leq \begin{cases} \left(\dfrac{\mu_1}{\mu_1+1}\right)^t & \text{if } t < \mu_1 \\[2ex] (\dfrac{\mu_1}{t-1})(\dfrac{t-1}{t})^t & \text{if } t \geq \mu_1 \ . \end{cases}$$

We observe that equality in (9) for $t < \mu_1$ is attained by the geometric distribution $p_k = \dfrac{1}{1+\mu_1}\left(\dfrac{\mu_1}{1+\mu_1}\right)^k$; $k = 0,1,2,\ldots$ and for

-18-

$t \geq \mu_1$ equality in (9) is attained by the geometric distribution G where $\overline{G}(k) = (\frac{\mu_1}{t-1})(\frac{t-1}{t})^k$; $k = 1,2,\ldots$. Hence, the inequalities given by (9) are sharp.

2.3 Comparison of the Continuous and Lattice Bounds:

Comparing Theorem 1 and Theorem 2, we remark that for μ_1 fixed, the continuous upper bound is everywhere smaller than the lattice upper bound. Similar comparisions hold for Theorem 5 and Theorem 6.

3. Application of Bounds

3.1 Motivation to Life Table Analysis:

We note that the failure rate function, $r(t)$ is familiar in other branches of science and is known by a variety of names. In the life sciences it is used by actuaries under the name of "Force of Mortality" to compute mortality tables [15]. The important work of life table analysis is to calculate survival rate, i.e., the proportion of persons surviving a specified interval of time. Various methods for calculating survival rate have been discussed in several research papers by Ederer [7]; Ederer, Cutler and Axtell [8], Berkson and Gage [4], etc. Some functions that have been used to define the force of mortality are Gompertz's formula and Makeham's formula. A detailed discussion concerning the estimation of the parameters in Makeham's law of mortality can be found in Grenander's paper [9].

Grenander remarks in his paper that we should specify carefully the probability assumptions to be used in mortality studies. The basic assumption is that our data forms a sample of independent observations of stochastic variables, with the same probability distribution except when sampling errors are serious. Further Grenander remarks in his paper that mortality estimates will be used to compute premiums, reserves and so on, and it is the future mortality that is of interest for making a prognosis for several decades forward in time. Since the development of the mortality depends upon factors that are very difficult to handle, it is likely that only a rough prognosis can be made.

We are interested in giving upper bounds on survival probability which will be useful in the life sciences in order to compare mortality rates in different populations. To give such bounds, which can be used for all ages, we need to have some a priori information about the failure rate function such as IFR or DFR, as we shall see in a cancer study.

3.2 Application of bounds to a cancer study:

Patient survival is generally accepted as the principal criterion for measuring the effectiveness of treatment in cancer. The American College of Surgeons requires the maintenance of a cancer registration and follow up program for approval of a hospital cancer program. We shall investigate a particular type of cancer from a large number of cancer studies given in [6] in order to apply the bounds.

We are interested in giving upper bounds for the survival probability (also called survival rate). Data of table 1 is taken from [6, p. 316], which represents the observed survival rates for breast and genito-urinary organs cancer for the period 1942-56 of both sexes. For this particular type of cancer there were 41,157 patients at the beginning of the year 1942. The observed survival rates were calculated by the usual formula and is shown in Table 1.

Our aim is to use this data to give an upper bound for the survival rate and to extend it for all ages.

TABLE - 1

Observed Survival Rates,
Breast and Genito-Urinary Organs Cancer,
Detailed Sites By All Ages, 1942 - 1956, Both Sexes.

Years after diagnosis. t	Observed Survival rates in percentage. $\hat{\bar{F}}(t)$	$Log_e\hat{\bar{F}}(t)$	Difference $Log_e\hat{\bar{F}}(t+1)-Log_e\hat{\bar{F}}(t)$
0	100.0	4.60517	- 0.33407
1	71.6	4.27110	- 0.18849
2	59.3	4.08261	- 0.13715
3	51.7	3.94546	- 0.11682
4	46.0	3.82864	- 0.09814
5	41.7	3.73050	- 0.09291
6	38.0	3.63759	- 0.08?24
7	35.0	3.55535	- 0.07719
8	32.4	3.47816	- 0.07696
9	30.0	3.40120	- 0.07616
10	27.8	3.32504	- 0.07080
11	25.9	3.25424	- 0.06789
12	24.2	3.18635	- 0.07283
13	22.5	3.11352	- 0.06425
14	21.1	3.04927	- 0.05354
15	20.0	2.99573	-------

Since successive differences of the function $Log_e\hat{\bar{F}}(t)$ are increasing in t except at $t = 12$, we can assume that the successive differences of the function $Log_e\bar{F}(t)$ are increasing in t for all values of t . The graph of the function $Log_e\hat{\bar{F}}(t)$ is shown in Figure 3.1.
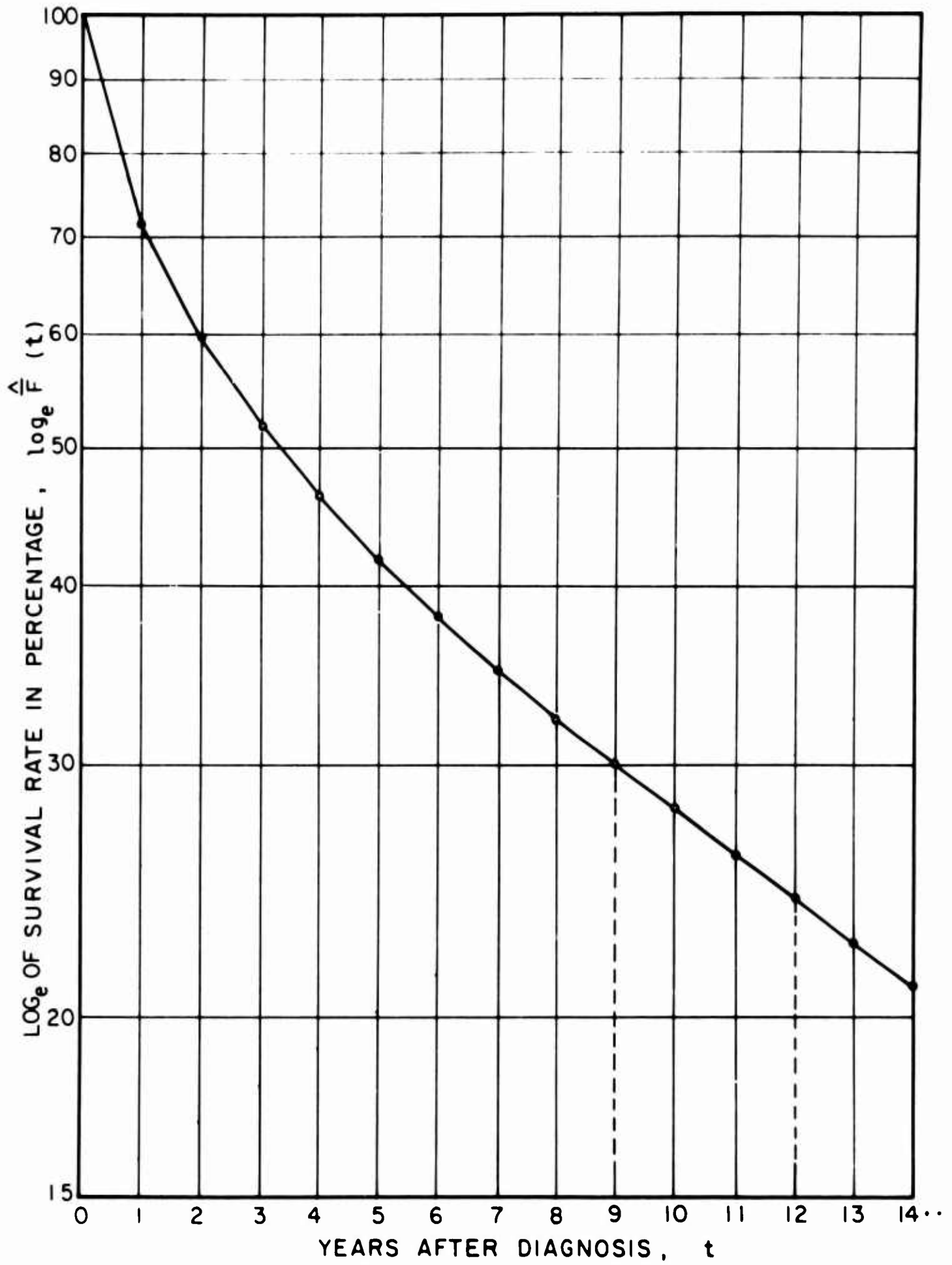
FIGURE 3.1

From successive differences of the function $\log_e \hat{\overline{F}}(t)$ and from the graph it is clear that $\log_e \overline{F}(t)$ is convex and therefore we claim that F is DFR.

Now we shall obtain upper bounds for 1-F using Theorem 5 and Theorem 6 and compare them.

We need to estimate μ_1 , the mean of F . Let $\hat{\mu}_1$ denote an estimate of μ_1 . Thus,

$$\hat{\mu}_1 = \sum_{t=1}^{\infty} \overline{F}(t)$$

$$= \sum_{t=1}^{14} \overline{F}(t) + \sum_{t > 14}^{\infty} \overline{F}(t)$$

$$= \mu_1' + \mu_1'' \text{ , say.}$$

From the data of Table 1, we have

$$\mu_1' = \sum_{t=1}^{14} \hat{\overline{F}}(t) = 527.2 \text{ .}$$

Hence, it remains to calculate μ_1'' . From the graph of the function $\log_e \hat{\overline{F}}(t)$ it is clear that for $t \geq 8$, $\overline{F}(t)$ has an exponential form and therefore, μ_1'' can be approximated by

$$\mu_1'' = \sum_{t > 14}^{\infty} \hat{\overline{F}}(t)$$

$$= \int_{14}^{\infty} g(x)dx \qquad \text{where } g(x) \text{ is an exponential curve.}$$

Let $g(x) = ce^{-a(x-14)}$ for $x \geq 14$.

So, $\log_e g(x) = \log_e c - a(x-14)$ implies $\log_e c$ is equal to the height at

-24-

$x = 14$ and $-a$ is the slope of the curve $\log_e g(x)$. This gives $c = 21.1$ and an estimate of $-a$ is given by

$$-a = \frac{\text{Log}_e \hat{\bar{F}}(9) - \log_e \hat{\bar{F}}(12)}{9 - 12} = -0.07162 .$$

Hence,

$$\mu_1'' = \int_{14}^{\infty} (21.1)e^{0.07162(x-14)} dx$$

$$= (21.1) \int_{0}^{\infty} e^{0.07162} dy$$

$$= 294.6 .$$

Finally, we obtain

$$\hat{\mu}_1 = 527.2 + 294.6$$

$$= 821.8$$

or $\qquad \hat{\mu}_1 = 8.218$ years .

Using an estimate of the mean of F as 8.218, we shall calculate upper bounds for $1-F$ in Table 2.

TABLE - 2

Upper Bounds on Survival Rate.

Years after diagnosis. t	Observed survival rates in %. $\hat{F}(t)$	Upper bounds on $\bar{F}(t)$ using Th. 6 in %.	Upper bounds on $\bar{F}(t)$ using Th. 5 in %.
0	100.0	100.00	100.00
1	71.6	89.15	88.54
2	59.3	79.47	78.37
3	51.7	70.85	69.42
4	46.0	63.16	61.47
5	41.7	56.31	54.42
6	38.0	50.20	48.19
7	35.0	44.75	42.66
8	32.4	39.89	37.78
8.218	31.9	38.94	36.79
(By Interpolation)			
9	30.0	35.55	33.58
10	27.8	31.80	30.23
11	25.9	28.79	27.48
12	24.2	26.29	25.18
13	22.5	24.16	23.24
14	21.1	22.39	21.58
15	20.0	20.82	20.13

From the last two columns we observe that the upper bounds on $\bar{F}(t)$ obtained by Theorem 6 are everywhere larger than those obtained by Theorem 5. These bounds can be compared graphically as shown in Figure 3.2 and Figure 3.3.
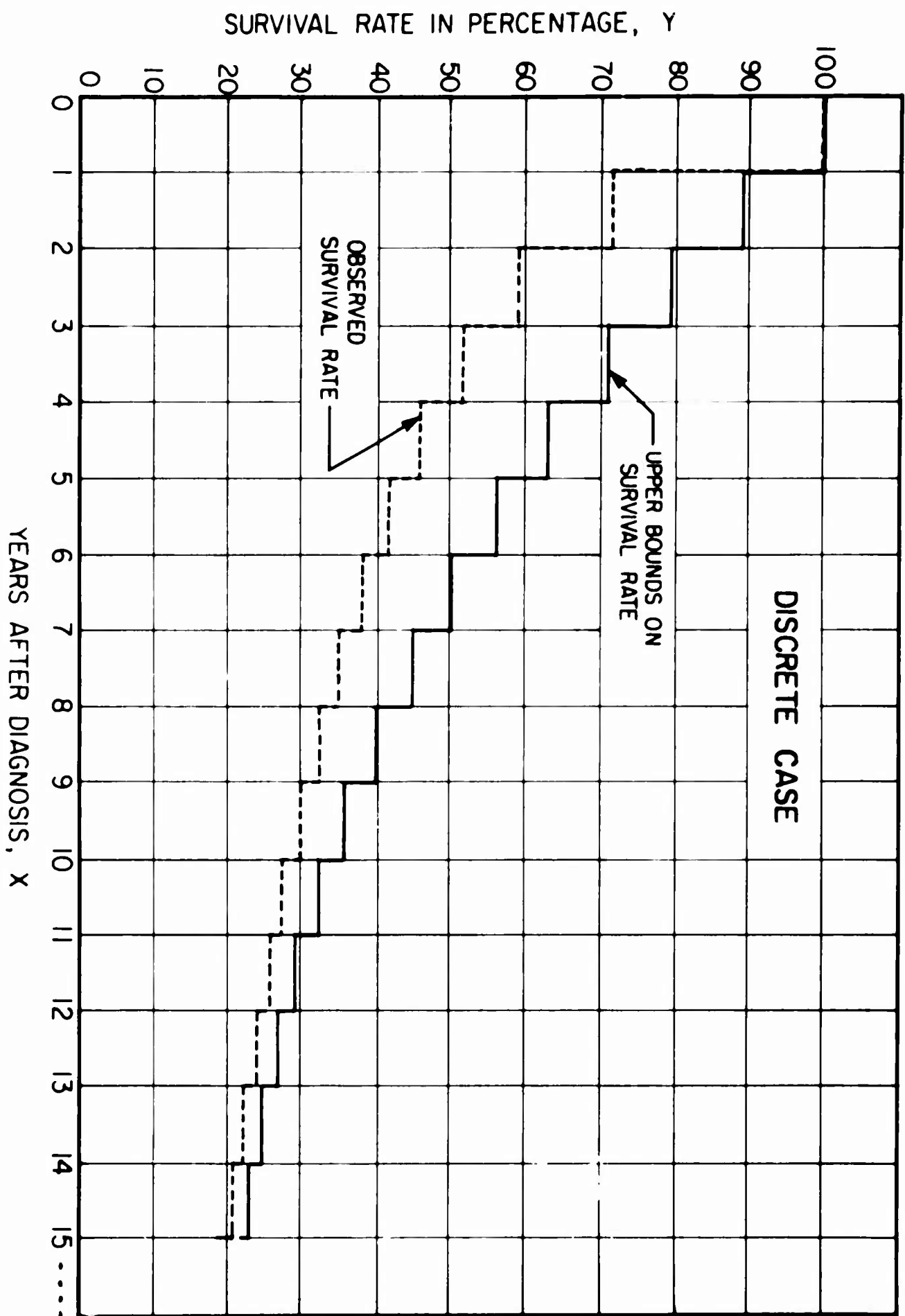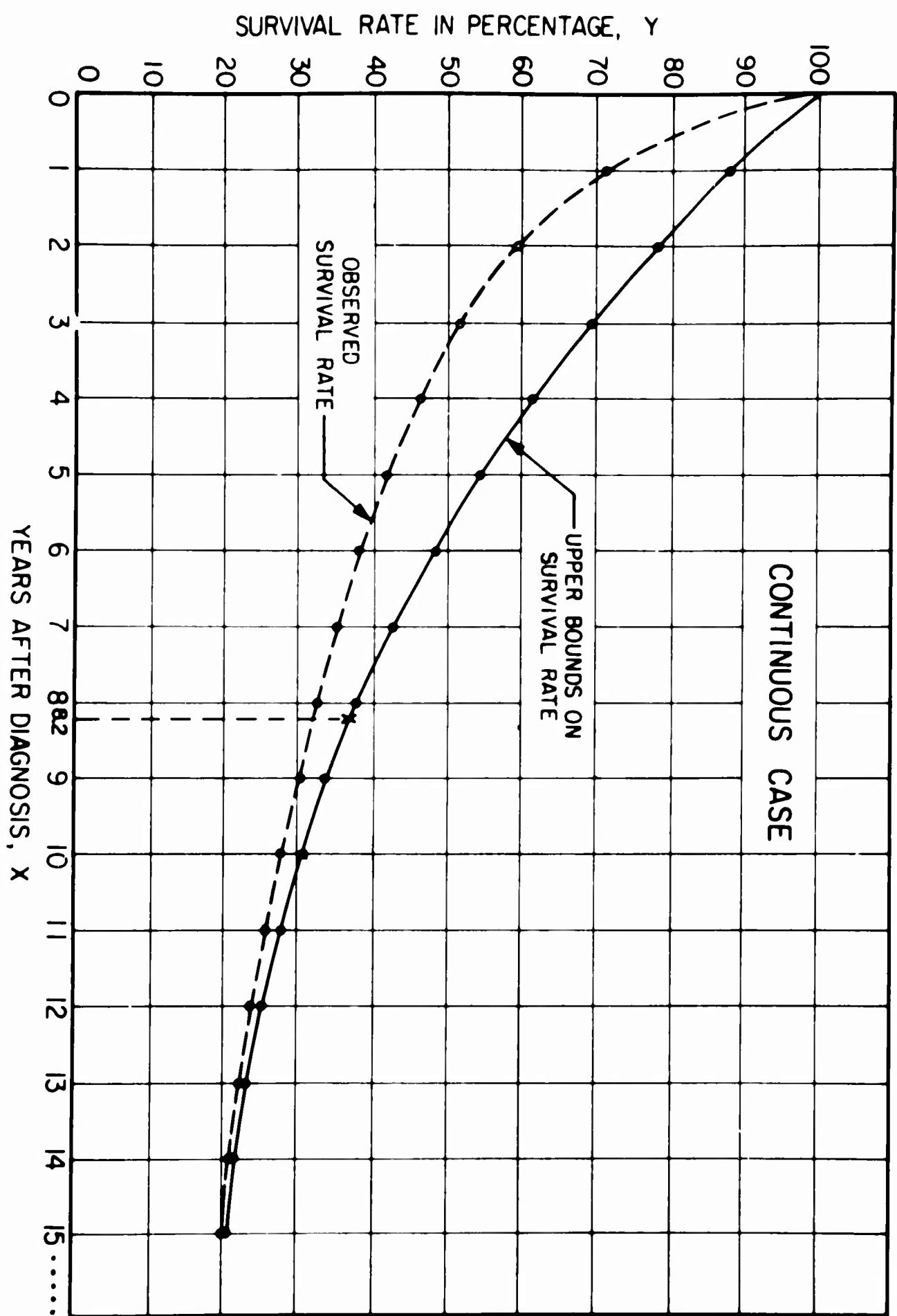
FIGURE 3.2

FIGURE 3.3

These results on bounds are of particular interest in comparing the mortality in different populations, when a priori information about the mortality function is known.

Suppose we are interested in comparing the survival rates of two types of cancer, say, stomach and lung. Let us assume that both have decreasing failure rate with mean μ_1 and μ_2 respectively. We note that one of the means is greater than the other say, $\mu_1 > \mu_2$. Also note that, the estimate of the means can be found by studying smaple information. Using Theorem 6, we can find upper bounds on survival rates for both stomach and lung cancer. Then, by comparing the upper bounds on stomach and lung cancer, we can estimate the greatest difference in survival rate for different periods.

3.3 Application of bounds to reliability theory:

In reliability theory, a decreasing failure rate would seem to correspond to some physical mechanism of improvement with age, so that the longer the unit survives the less the chance of failure in the next unit of time. We will discuss the reliability problem given in [13] by establishing the upper bound.

Proschan was interested in obtaining information as to the distribution of failure intervals for the air conditioning system of each member of a fleet of Boeing 720 jet airplanes [13]. Records were kept for the time of successive failures (to the nearest hour) of the air conditioning system [13, p. 316].

First we note that it was proved in [13] that the failure distribution formed by pooling failure intervals for a given component operating in different types of equipment would be DFR. Thus, we know that the data follows a DFR distribution and hence we will give an

-29-

upper bound on survival probability using Theorem 6. We remark that
an estimate of the parent mean can be estimated by sample information
and it was found in [13] that the mean of all the failure intervals
was 93.14. Since the sample size is large, this estimate should be
fairly accurate. Using 93.14 as an estimate of the parent mean,
we give theoretical upper bound for the pooled data.

Figure 3.4 shows the upper bounds on survival probability.
The cross-over in the beginning for the observed survival probability
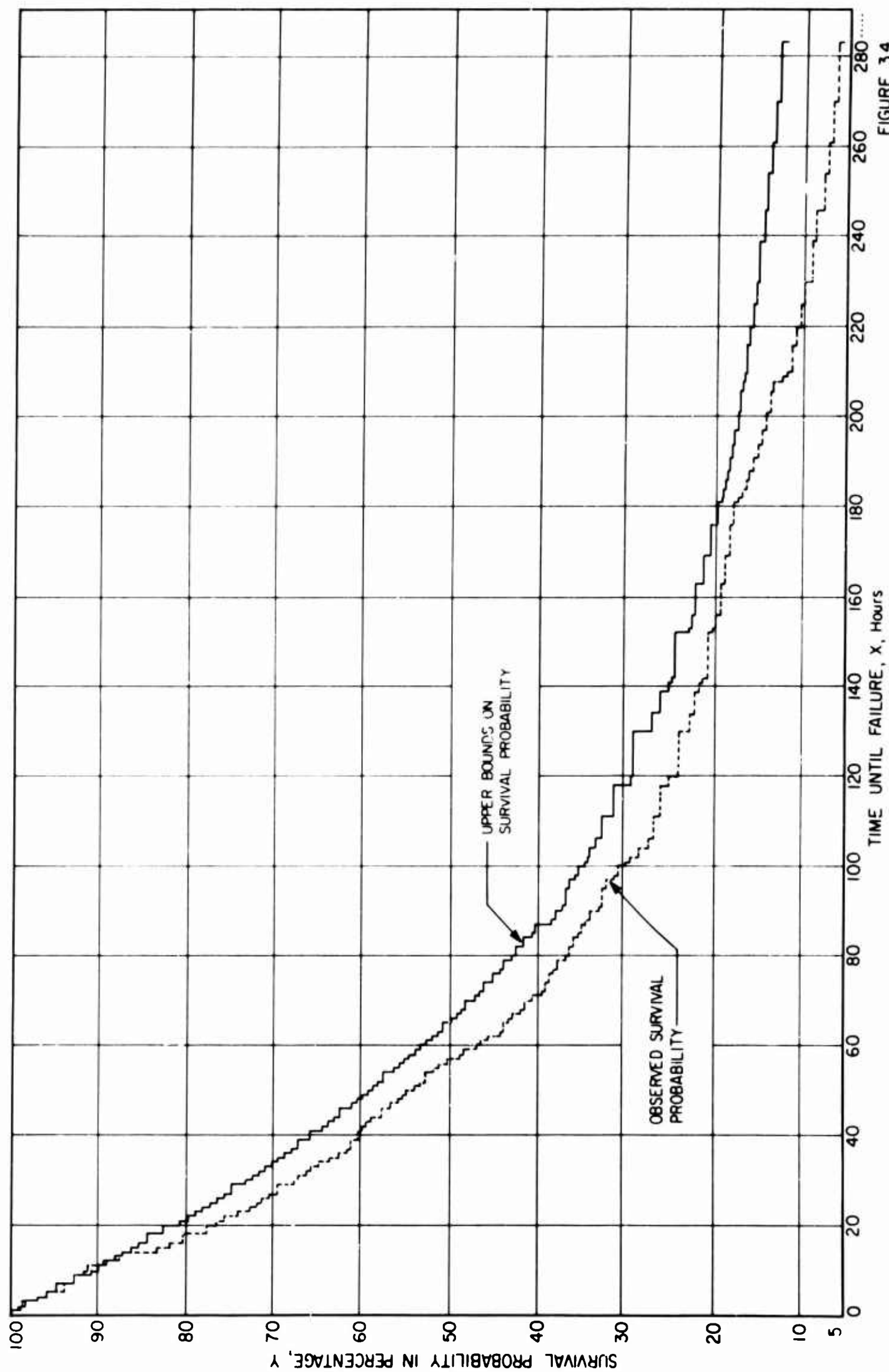may be due to random fluctuation.

FIGURE 3.4

UPPER BOUNDS ON
SURVIVAL PROBABILITY

OBSERVED SURVIVAL
PROBABILITY

TIME UNTIL FAILURE, X, Hours

SURVIVAL PROBABILITY IN PERCENTAGE, Y

31

Acknowledgement

The author would like to gratefully acknowledge the assistance
and encouragement of his thesis advisor Richard Barlow and also
Dr. Albert Marshall of the Boeing Scientific Research Laboratories.

Bibliography

[1] Barlow, R.E., and Proschan, F., Mathematical Theory of Reliability. New York: John Wiley and Sons, Inc., 1965.

[2] Barlow, R.E., and Marshall, A.W., "Bounds for distributions with monotone hazard rate, I," Annals of Mathematical Statistics, 35 (1964), 1234-57.

[3] Barlow, R.E., Marshall, A.W., and Proschan, F., "Properties of probability distributions with monotone hazard rate," Annals of Mathematical Statistics, 34 (1963), 375-89.

[4] Berkson, J., and Gage, R.P., "Specific Methods of Calculating Survival Rates of Patients with Cancer," Chapter 37 in Treatment of Cancer and Allied Diseases by Pack, G.T., and Ariel, I.M., Vol. I. New York: Paul B. Hoeber, Inc., 1958.

[5] Birnbaum, Z.W., and Saunders, S.C., "A statistical model for life-length of materials," Journal of the American Statistical Association, 53 (1958), 151-160.

[6] "Cancer Registration and Survival in California- California Tumar Registry" published by State of California, Department of Public Health, Berkeley, California, 1963.

[7] Ederer, F., "A Parametric Estimate of the Standard Error of the Survival Rate," Journal of the American Statistical Association, 56 (1961), p. 111.

[8]   Ederer, F., Axtell, L.M., and Cutler, S.J., "The Relative Survival Rate," National Cancer Institute, Bethesda, Maryland, Monograph No. 6.

[9]   Grenander, Ulf, "On the Theory of Mortality Measurement," Skandinavisk Aktuarietidskrift, (1956), 1-55.

[10] Karlin, Samuel, Total Positivity and Applications to Probability and Statistics. Stanford University Press, Stanford, California, (in preparation).

[11] Karlin, Samuel, and Shapley, L.S., "Geometry of Moment Spaces," Memoirs of the American Mathematical Society, No. 12, 1953.

[12] Karlin, Samuel and Studden, Forthcoming book on Chebyshev Systems and Inequalities, (1965).

[13] Proschan, Frank, "Theoretical explanation of observed decreasing failure rate," Technometrics, 5 (1963), 375-83.

[14] Spurgeon, E.F., Life Contingency. Published for the Institute of Actuaries, University press, Cambridge, England (1947).

[15] Steffensen, J.F., Some Recent Researches in the Theory of Statistics and Acturial Sciences. Cambridge University press, England (1950).